DOCUMENT RESUME

ED 057 161 VT 010 835

AUTHOR Johnson, Kirk A.
TITLE Retention of Electronic Fundamentals: Differences Among Topics.
INSTITUTION Naval Personnel Research Activity, San Diego, Calif. Navy Training Research Lab.
SPONS AGENCY Office of Naval Research, Washington, D.C.
REPORT NO TB-STB-70-71
PUB DATE Aug 69
NOTE 30p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Charts; *Criterion Referenced Tests; *Electronics; *Enlisted Men; *Retention Studies; *Student Evaluation; Tables (Data); Tests; *Training

ABSTRACT

Criterion-referenced tests were used to measure the learning and retention of a sample of material taught by means of programed instruction in the Avionics Fundamentals Course, Class A. It was found that the students knew about 30 percent of the material before reading the programs, that mastery rose to a very high level on the immediate posttest, and that about half of the improvement was lost by the end of the course (an interval of about 96 days). There was considerable variation in item difficulty by the end of the course. Most of this variation was independent of topic difficulty or measures of time difficulty obtained from the early posttests. Instructors (who were also experienced technicians) were asked to indicate the items that were most relevant to subsequent instruction or to performance on the job. These ratings were not very reliable. The indicated items did not differ appreciably from the remaining items in terms of student proficiency. It was concluded that if the instructors were correct in their ratings, there was enough forgetting to hinder a number of students in learning from subsequent courses and in performing their assigned duties on the job. (Author)

106

ED0 57161

AD

U. S. NAVAL PERSONNEL RESEARCH ACTIVITY

SAN DIEGO, CALIFORNIA 92152

TECHNICAL BULLETIN STB 70-1

AUGUST 1969

RETENTION OF ELECTRONIC FUNDAMENTALS: DIFFERENCES AMONG TOPICS

Kirk A. Johnson

NPRA

NAVAL PERSONNEL RESEARCH ACTIVITY SAN DIEGO, CALIFORNIA

AN ACTIVITY OF THE BUREAU OF NAVAL PERSONNEL

1

NOTE: The contents of this publication do not necessarily represent the official position or policy of the Department of the Navy.

AD

RETENTION OF ELECTRONIC FUNDAMENTALS: DIFFERENCES AMONG TOPICS

by

Kirk A. Johnson

August 1969

PF39.522.004.01.42
Technical Bulletin STB 70-1

E. I. Jones, Ph.D., Director, Navy Training Research Laboratory
E. E. Dudek, Ph.D., Technical Director
K. E. KUEHNER, Commander, USN
Commanding Officer

Navy Training Research Laboratory
Naval Personnel Research Activity
San Diego, California

SUMMARY AND CONCLUSIONS

Problem

There are several points within a training sequence at which it would be helpful to have an absolute measure of student proficiency. A measure of proficiency before training would indicate the topics on which more or less training would be required. In an operational system it might identify the students who are in need of remedial instruction. A measure at the completion of training would indicate the parts of the course that were in need of revision. A measure at the point of application would indicate the topics that were in need of review. These measures cannot be derived from conventional norm-referenced tests of the kind used in most schools; they require, instead, special criterion-referenced tests.

Background

The purpose of the present investigation was to explore some of the benefits and problems associated with the use of criterion-referenced tests in an operational training situation. The study was based on the material being taught in the first phase of the Avionics Fundamentals Course, Class A. All materials were originally taught by means of programmed instructional booklets.

Approach

Proficiency was measured by means of the criterion-referenced tests that had been used in the validation of the programs. Measures were obtained on a pre-test, an immediate post-test, and at intervals of one day, seven days, 28 days, 96 days, and three years following original learning. All except the last measure were obtained from the same students, but the particular items were counterbalanced so that no student would take a given item twice.

Findings, Conclusions, Recommendations

It was found that students knew about 30% of the material before they read the programs, that proficiency reached a very high level on the immediate post-test, but that about half of this improvement was lost by the end of the course. Over several years on the job, proficiency dropped to a level that was less than 20 percentage points above that found on the pre-test. By the end of the course there were large differences in proficiency on different items (S.D. = 27 percentage points). These were reliable differences, but little of the variance could be attributed either to general topics or to measures of item difficulty obtained at the earlier testing points.

Instructor ratings were used in an effort to identify the items which covered knowledge that would be most relevant to performance on the job or to learning in subsequent courses. Most of these ratings were fairly unreliable. In any case, the items identified by means of these ratings did not differ appreciably from the remaining items in terms of student proficiency.

If the instructor's opinions can be taken at face value, then the forgetting found in this study would be enough to hinder many students in their learning of the material taught in subsequent courses and in their performance on the job.

Information of the kind found in this study should be very helpful in the design and control of a training sequence, but the systematic collection of such information, particularly for a lengthy training sequence, would be a major undertaking.

REPORT USE AND EVALUATION

Feedback from consumers is a vital element in improving products so that they better respond to specific needs. To assist the Chief of Naval Personnel in future planning, it is requested that the use and evaluation form on the reverse of this page be completed and returned. The page is preaddressed and franked; fold in thirds, seal with tape, and mail.

Department of the Navy

Postage and Fees Paid
Navy Department

Official Business

Chief of Naval Personnel (Pers-A3)
Department of the Navy
Washington, D. C. 20370

Report Title & No: RETENTION OF ELECTRONIC FUNDAMENTALS: DIFFERENCES AMONG TOPICS (STB 70-1)

1. Evaluation of Report. Please check appropriate column.

FACTORS	RATING			COMMENTS
	LOW	AVE	HIGH	
Usefulness of Data				
Timeliness				
Completeness				
Technical Accuracy				
Validity of Recommendations				
Soundness of Approach				
Presentation and Style				
Other				

2. Use of Report. Please fill in answers as appropriate.

a. What are your main uses for the material contained in the report?

b. What changes would you recommend in report format to make it more useful?

c. What types of research would be most useful to you for the Chief of Naval Personnel to conduct?

d. Do you wish to remain on our distribution list?

e. Please make any general comments you feel would be helpful to us in planning our research program.

NAME: ______________________ CODE: __________

ORGANIZATION: ______________________

ADDRESS: ______________________

CONTENTS

TABLES

FIGURES

RETENTION OF ELECTRONIC FUNDAMENTALS: DIFFERENCES AMONG TOPICS

A. Introduction

The training of an electronics technician usually begins with a unit of instruction on electronic fundamentals. The purpose of this unit is two-fold: first, to provide a "vocabulary" of concepts that can be used to simplify and facilitate subsequent instruction and, second, to provide a set of principles that can be used as a basis for the deduction of proper maintenance procedures in those situations which have not been covered by specific instructions.

If a system of this kind is to work effectively, it is obvious that the material taught in the initial unit must be remembered long enough for it to play its intended role in the system. Some of it need be retained only to the point at which it can provide the basis for subsequent training; the remainder must be retained until it can be applied on the job.

There have been several studies in which the retention of these basic concepts and skills was measured at various points within the training sequence and at various intervals after the completion of formal training. Most of these studies were done at a time before a clear distinction had been drawn between criterion-referenced tests, which are designed to provide an absolute index of mastery, and norm-referenced tests, which are designed to provide an index of the relative differences between people. (Glaser, 1963). Since a large number of multiple-choice items were readily available from the norm-referenced tests used during training, it was only natural that these tests be used as the primary source of items for the tests of retention. These studies reveal a decline in the mastery of basic concepts and skills that begins during training and continues out onto the job, but, because of the ambiguities inherent in the use of a norm-referenced test as an absolute index of mastery, there has been no way to estimate the practical significance of this decline. The primary purpose of the present study was to investigate the retention of this material through the use of criterion-referenced tests.

B. Procedures

1. Training Content

At the time of this study the Avionics Fundamentals (AFU) course, taught at the Naval Air Technical Training Center, Memphis, lasted 16 weeks. For most students, this course was followed by an eight to 12 week course on a particular class of devices (radars, fire-control systems, etc.), and this in turn was followed by various courses on the equipment used in specific weapons systems. It would be helpful to have information on the retention of materials taught throughout this sequence, but because of practical constraints, the present study was limited to the first phase of the AFU course. This phase lasted five weeks and provided an introduction to basic a-c and d-c theory.

The topics were further restricted to those that had been taught by means of programmed instruction. At the time of this study, approximately 40% of the five week period was devoted to in-class use of programmed instruction, approximately 20% was devoted to laboratory work, and the remaining 40% was devoted to lectures, discussions, demonstrations, drills, reviews, tests, etc. This restriction was prompted by the fact that the tests used in the validation of the programmed booklets provided "exhaustive" criterion-referenced tests over all training objectives covered by the programs. The effect of this restriction was probably fairly slight. The programmed booklets had been designed to provide a reasonably self-sufficient introduction to a-c and d-c theory. It is doubtful that this particular mode of instruction had a major effect on the retention data, either. Much of the material covered by the programs received further elaboration through other media of instruction. In addition, previous studies with these same materials and tests (Mayo and Longo, 1964; Longo and Mayo, 1965) have shown that after a period of several days the retention of students taught by means of programmed booklets is similar to the retention of students taught by more conventional means.

The final restriction was prompted by the sheer quantity of material that remained. In order to reduce this material to a more manageable level, without at the same time changing its quality, all even-numbered programs were excluded from the study. This left a total of 19 programs that covered 215 training objectives. A list of these programs can be found in Appendix A.

2. Students

Most of the data were collected from 85 students in a single class who were present at each of the major testing points during the course. These were students who proceeded through the course at a normal rate, without setbacks because of academic deficiencies or accelerations because of academic superiority. The original class contained 141 students. Of these, 9 were dropped for academic reasons, 7 were removed for administrative reasons, 21 were set back to later classes for academic reasons, 17 were placed in accelerated sections, and 2 were simply absent at one of the major testing points. Since a normal graduating class would contain in addition to the students who convened with that same class, other students who, because of setbacks or accelerations, convened with other classes, the 85 students used in this study were not strictly representative of the normal school output. The number of graduates lost because of superior performance (17) was fairly close to the number of graduates lost because of inferior performance (21), however, so the average performance of this group is probably fairly similar to that of a normal group, even though the variance of its performance is probably smaller.

Some additional data were collected from a group of 29 technicians who had just returned from the fleet in order to attend an advanced course in avionics (AVIB). All of these technicians had attended the AFU school at some time in the past. The median interval between graduation from the AFU school and testing was about three years. There are a variety of selective factors that might have affected the quality of this second sample, but the available evidence indicates an overall effect that is fairly small. Table 1 contains background data on both samples.

TABLE 1

Background Data on AFU and AVIB Samples

Sample	ETST	GCT	ARI	AFU Final Average
AFU	65.2	64.4	61.5	80.3
AVIB	63.9	61.0	61.3	78.2
Diff	1.3	3.4	.2	2.1

3. Test Items

The retention tests, as was mentioned earlier, were made up of the criterion tests that had been used to validate the programs. A given training objective was generally covered by a single item, though many of these items were actually a composite of several fairly distinct problems. A student might, for example, be required to calculate several circuit values, or to transform several values from one set of units to another. Most of the problems required a written answer of some kind, though a few of them were in a multiple-choice or matching format. Some of the original items were modified slightly in order to clarify their meaning when viewed in isolation. The specific values used in the problems were changed whenever this could be done without an obvious effect on the difficulty or the essential content of the item, since it was feared that without such changes the students might tend to respond on the basis of rote memory instead of the intended calculations.

4. Testing Plan

Each of the criterion tests covered, on the average, 11.5 training objectives. Each of these tests was broken down into six sets of items that were as closely matched as possible. For the single test that contained fewer than six items, it was possible to split composite items so as to form the required number of sets.

The initial class of 141 students was broken down alphabetically into six groups of roughly the same size. These initial groupings were retained throughout the course, but losses of various kinds reduced their sizes. By the end of the course, there were from 12 to 17 students in each of the six groups.

The items associated with a given program were given to students at each of six points. The first of these was immediately prior to the administration of the program. The second was at the end of the classroom time allotted to the program. The next three were 1 day, 7 days, and 28 days after the completion of the program, respectively. The final point was at the end of the course, which was, on the average, 96 days after the completion of the program.

For each program, the six item sets, six student groups, and six testing points formed a Latin square. Each of the item sets and student groups was represented once and only once at each of the testing points, and no student repeated an item he had encountered previously.

Since there was no problem of retesting with the AVIB students, each student was given half the item sets. This provided data on from 13 to 15 students per item at this point, roughly the same amount of data that was available at each of the testing points used with the AFU students.

5. Test Scoring

For many of the items the answers were clearly right or wrong. For others, however, there were varying degrees of "rightness." In most cases, the scoring standards were fairly lenient. Partial credits were given when a student missed only part of a composite item or when he indicated a substantial knowledge of the correct response. They were given on computational problems, for example, when the error could be traced to a simple mistake in arithmetic or in the designation of units, even though errors of this kind could be quite serious on the job.

All items were scored by a single individual. For the AFU students, all responses to a given item were scored at the same time, without knowledge of the student group or the testing point from which a given response had been obtained. The items from the AVIB sample were scored at a later date, but against a background provided by the items from the AFU sample. Most of the responses used by the AVIB sample had also been used by the AFU sample, so there was little room for bias.

C. Results

The results will be broken down into three sections. The first describes average student performance at each of the seven testing points. The second describes an attempt to draw finer distinctions among the items in terms of the uses that will be made of the knowledge or skill being tested. The final section describes the relationships between performance on various items or topics at different points following the initial instruction.

1. Overall Retention

Since all six student groups from the AFU course took all 218 items, it was possible to compute a score for each student by averaging across all items. The averages of these individual scores for the six student groups were .69, .69, .71, .69, .69, and .69, a remarkably close agreement.

Since the groups were so evenly matched, it was possible to combine them for an item by item summary of performance at each of the testing points. This has been done in Figure 1. In discussing these data it will be convenient to treat the ordinate as if it represents the percentage of students who were completely right on the items in question, even though, because of the partial credits, this is not strictly the case.

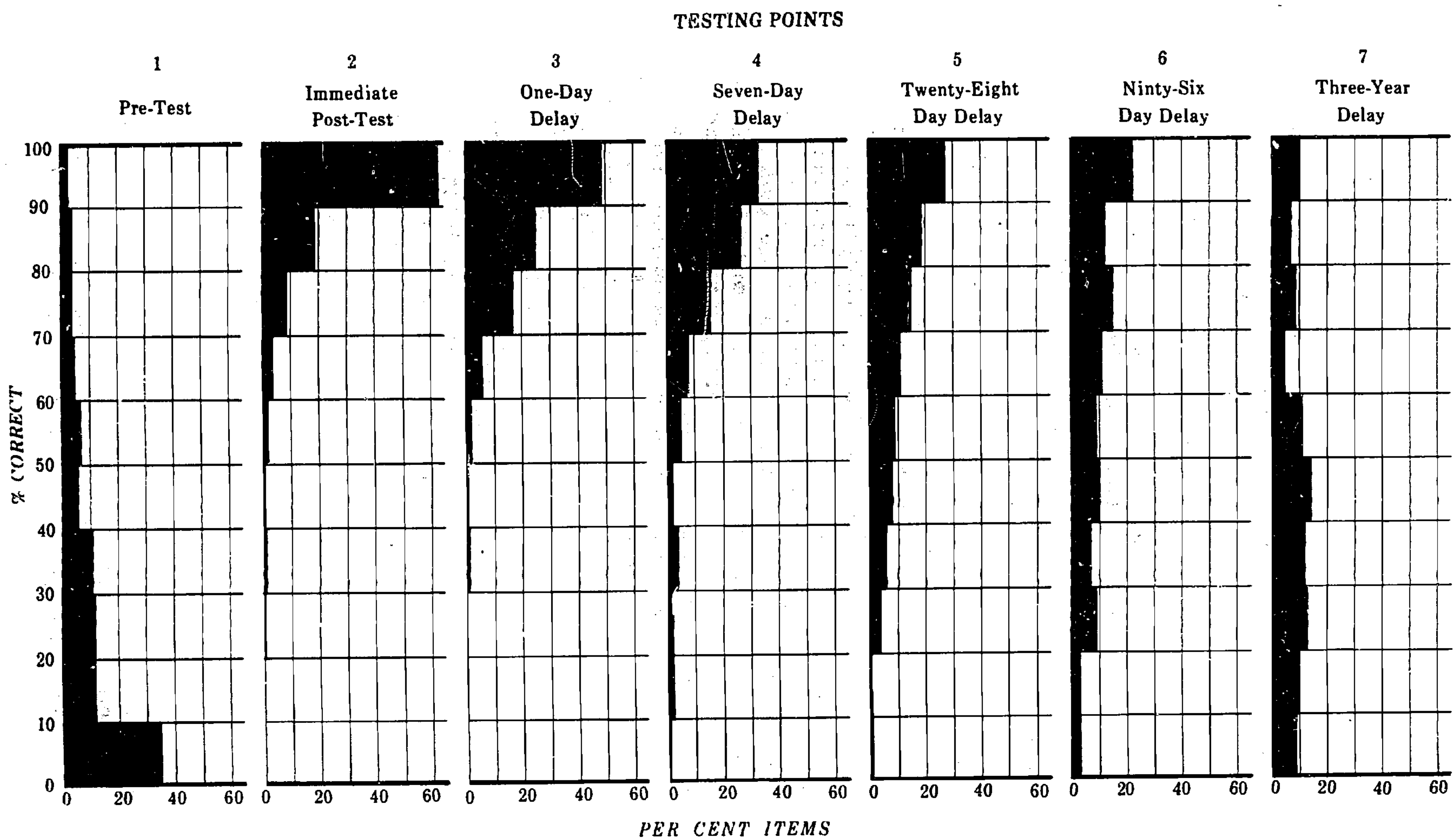

Figure 1. Distributions of Item Difficulties at Seven Testing Points.

An examination of the first testing point indicates that there was a moderate amount of this material known by the students before they took the programs. If it could be assumed that equal amounts of time were spent on each objective, then almost 30% of the total training time would have been spent in teaching students things that they already knew. There are several reasons for rejecting this assumption, however. First, there was probably less material devoted to the objectives with which the student might be expected to have some familiarity. Second, branches that would permit the student to bypass the material that he already knew were available in several of the programs. Finally, the student could adjust his own pace to some extent by skimming over the material with which he was already familiar and slowing down for the material that was new.

It should be remembered that the data at this point represent the students' knowledge immediately before taking the various programs and not their knowledge at the beginning of the course. Many objectives were covered in home study assignments, and some objectives in the later programs were introduced by programs earlier in the sequence.

The data of the second testing point indicate a high degree of mastery, with roughly half the items being answered without a single error. The average, over all items, is 90%. Even so, this is an underestimate of the actual maximum, since some of the students had not completed the programs at this time (those who did not finish a program within the allotted time finished it during later class periods or in the evening after school). In fact, as a result of these non-completions, only 58% of the students demonstrated mastery of 90% of the items tested at the second testing point, whereas, more than 90% of the students used in the original validation of the programs, all of whom were allowed to complete the programs, reached this same criterion. If the students in the present study had been tested at the actual completion of the programs, their average score would probably have been several percentage points higher than that which was found at the second testing point.

Following the second testing point the highly skewed distribution of items becomes flatter and flatter, until, by the seventh testing point, it is an almost rectangular distribution extending from 0% to 100%. If the midpoints of these distributions were plotted against time, they would form a negatively accelerated function that resembles the classic curve of forgetting. It should be remembered, however, that these data were gathered from a task that differs considerably from the usual laboratory task. The intervals are filled with learning activities that provide massive opportunities for both positive and negative transfer. For many of the items there is a great deal of direct rehearsal and practice.

By the end of the course, the scores of the AFU students have dropped about half the distance between their highest level and their original level. The scores made by the AVIB students are less than 20 percentage points above the scores made by the AFU students on the pre-test. An examination of the correlations between student characteristics and proficiency indicated that none of the adjustments for differences between the two samples would increase the scores at the seventh testing point by more than 1.8 percentage points.

2. Specific Retention

There is obviously a considerable amount of forgetting that takes place by the end of the course, but its practical importance cannot be determined without a consideration of the individual items involved. A good deal of the material taught in this initial segment is taught purely as an aid to the learning of additional material. If this additional learning takes place within the AFU course itself, then there is no reason why the original material cannot be forgotten without any real loss to the student.

In order to identify the material that should not be forgotten, the various item sets were submitted to instructors from the AFU course, who were asked to indicate, for each item, whether the knowledge covered would be needed on the job and/or in the various courses which follow the AFU course. A copy of these instructions can be found in Appendix B. The instructors had all been through the training sequence at some time in the past and had served at least one tour of duty in operating units. Each of the six item sets was evaluated by nine instructors, but, in general, they were not very reliable. A summary of these ratings can be found in Appendix C.

The most reliable index was obtained from the ratings of relevance to the job. The instructors endorsed about 43% of the items, the average correlation between instructors was .31, and the overall reliability was .81. The items that had been endorsed by at least seven of the nine instructors were selected for further analysis. There were 44 such items, representing about 20% of the total pool. Performance on these items has been graphed in Figure 2. It can be seen that the students were somewhat more proficient on these items than they were on the remaining items. By the seventh testing point, the difference had increased to about 20 percentage points. This difference provides some substantiation for the instructors' judgements, since one would expect that the use of this information on the job would maintain proficiency at a level above that found for the less frequently used information. The absolute level of proficiency, however, is rather disappointing. If this information is "really needed on the job," as was stated by the instructors, then it would not seem unreasonable to expect a level of proficiency approaching 100%; instead, it was found to be 65%.

Since the instructors' ratings may have been based on the assumption that "the job" would entail maintenance at the level of individual components, part of the deficiencies noted above might be attributed to the fact that many of the students in the AVIB sample had little or no experience on this kind of job. In order to check on this possibility, the AVIB school sample was divided into two groups of roughly the same size. The first group had had more than six months experience on jobs that required maintenance at the level of individual components. Most of this experience was in Intermediate Maintenance Activities. The second group had had six months or less experience in this type of work. A good deal of their time had been spent in Organizational Maintenance, though some technicians worked at jobs that were almost completely divorced from normal maintenance activities.

The two groups differed by less than three percentage points on total scores. They differed by less than seven percentage points on the "needed" items, a somewhat larger difference, but hardly one that could be used to ...ain the deficiencies noted earlier.

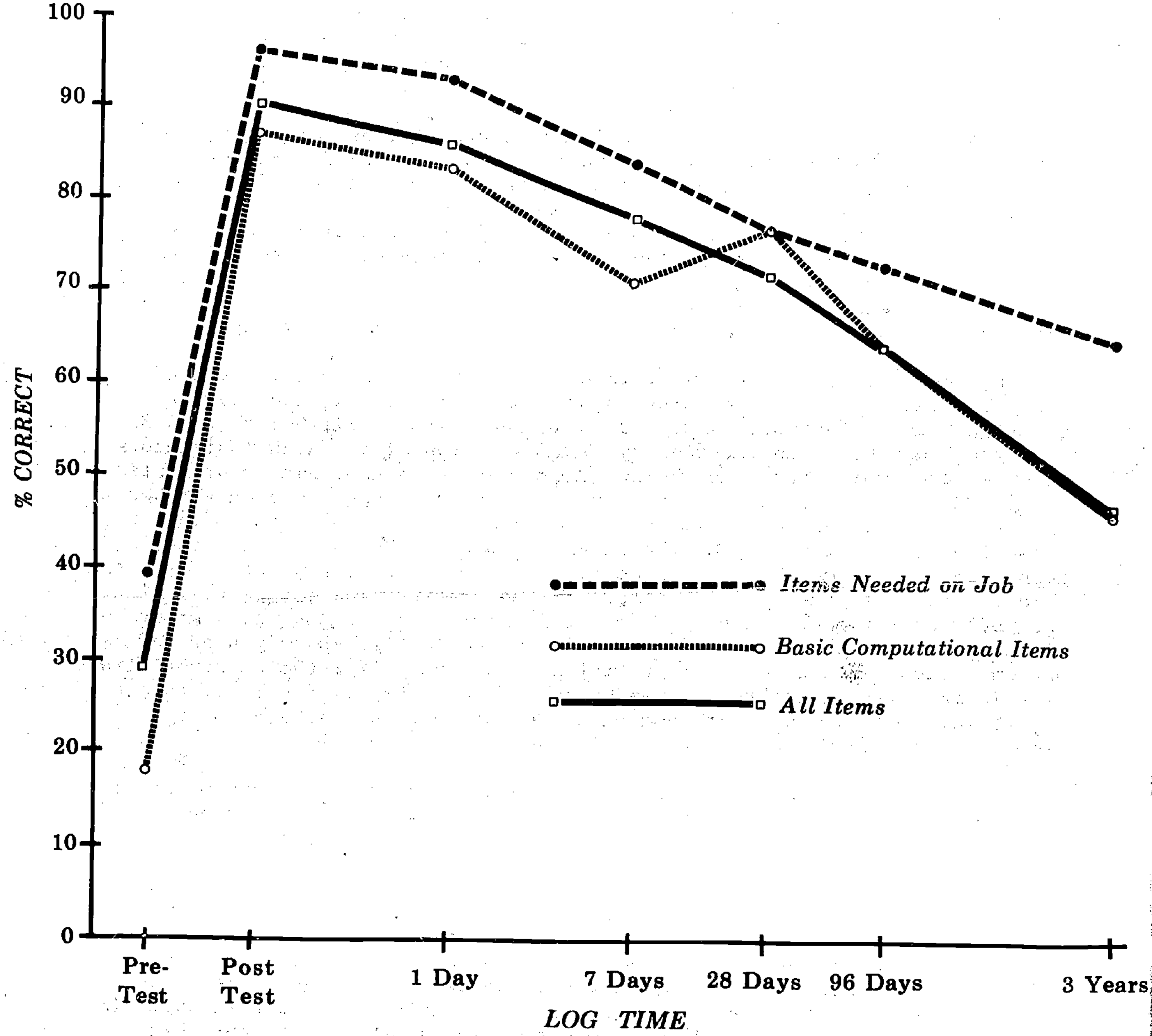

Figure 2. Performance on Three Kinds of Items at Seven Testing Points.

A second analysis was made of the 60 items that had been endorsed by at least seven of the nine instructors as being required for subsequent training. The average instructor endorsed 59% of the items, the average correlation between instructors was .12, and the overall reliability was .53. Performance on these items was quite similar to performance on the remaining items. In fact, mean performance on these items was from one to three percentage points below mean performance for all items at each of the seven testing points.

Another analysis was performed on a set of 13 computational items from the areas indicated in Table 2. It was felt that these items would provide a fairly objective benchmark for the remaining items, since the content of computational items can be readily specified and both the questions and answers are free from the ambiguities encountered in some of the more purely verbal items. These particular items were selected from a larger set of computational items because the skills involved, from an admittedly subjective point of view, appeared to be the most basic.

TABLE 2

Areas Covered by Computational Items

Area Tested	Number of Items
Conversions from one metric prefix to another (e.g., 20 ma = ____ μa)	2
Solution of simple problems using metric prefixes (e.g., 100 kv ÷ 10 ma = 10 ____ Ω)	2
Calculation of voltage, current, power, or resistance in various parts of simple (2 to 3 resistors) resistive parallel circuits (e.g., if total current is 3a, and both R_1 and R_2 are 20 Ω, what is the total power being dissipated?)	6
Calculation of voltage, current, or impedance transmitted across a transformer (e.g., with 60 turns in primary, 100 turns in secondary, and 120 v applied to primary, what voltage is induced in the secondary?)	3

As can be seen in Figure 2, performance on these items was also quite similar to performance on the remaining items. The scores of the AVIB students who were more experienced in component level maintenance exceeded those of the less experienced students by about 13%. A more detailed discussion of computational skills can be found in Appendix D.

The last of the separate analyses was made on a group of ten composite items, all of which required the technician to indicate the effect (i.e., increase, decrease, or remain the same) that certain changes in simple RCL circuits (i.e. an increase or decrease in capacitance, inductance, resistance, or frequency) would have on certain aspects of circuit operation (e.g., X_L, Z_t, I_C, E_R). It was found that the average score at the sixth testing point was 63% and that the average score at the seventh testing point was 33%. The latter is exactly what would have been expected on the basis of guessing alone.

3. The Measurement of Mastery

Much of the recent interest in criterion-referenced tests has centered on their use as a means for providing quality control in instructional systems. It has been recognized that certain of the traditional psychometric considerations are not relevant in such applications, but there is very little data available on the characteristics that are relevant.

The scores from the current study were analyzed so as to provide information on the reliability of the criterion-referenced tests and on the extent to which tests administered at various points in time agree with one another. Reliability, as used here, refers to the tests' ability to make reliable discriminations among various training objectives or lessons, rather than to their ability to make reliable discrimination among students. Similarly, agreement across testing points is measured in terms of items or topics rather than students.

a. Single Training Objectives. The reliability of the tests at each of the testing points was estimated separately for each of the student groups. In each case the estimate was based on a score matrix that was approximately 36 (items) by 14 (students). The final estimate for a given point was then calculated by taking a weighted average, using Fisher's z transformation, over the six individual estimates. These final estimates have been placed, in parentheses, along the principal diagonal of Table 3. The variations in reliability follow fairly closely the variations in the standard deviations of item scores, which can be found in the last column of the table.

The correlations between testing points were based on the average item scores at each point; there was no division into student groups. Such a procedure will treat any differences between student groups as error, but as was stated earlier, there was very little difference between the six groups from the AFU school. A similar check could not be made on the six groups from the AVIB school, so it was simply assumed that they, too, would be fairly similar to one another. These correlations can be found above the principal diagonal in Table 3.

Since there were not very many students per item, a deficiency that could be readily corrected with additional testing, the correlations between test points were corrected for attenuation. The corrected coefficients can be found below the principal diagonal in Table 3.

If the first testing point is excluded, the rest of the matrix falls into a generally simplicial form. Tests that are close to one another in time are more highly correlated than those which are more widely separated in time. The correlations with the pre-test, on the other hand, tend to increase with

TABLE 3

Reliabilities and Intercorrelations for
Seven Testing Points: Objectives

Testing Points	Testing Points 1	2	3	4	5	6	7	S.D.
1	(.90)	.21	.32	.32	.40	.44	.54	.27
2	.27	(.69)	.60	.40	.39	.32	.31	.13
3	.41	.89	(.66)	.59	.53	.47	.42	.13
4	.37	.53	.79	(.85)	.67	.60	.53	.22
5	.46	.51	.70	.79	(.85)	.74	.64	.23
6	.49	.40	.61	.69	.85	(.89)	.73	.27
7	.61	.40	.55	.61	.74	.82	(.89)	.28

increasing separations from the point of original learning. Both patterns hold for the corrected matrix as well as for the original matrix.

The most important finding, from a practical point of view, is that performance on the early post-tests does not provide a very powerful basis for predicting performance at the more delayed testing points. The highest of these correlations, that between testing point 3 and testing point 6, accounts for only 37% of the variance in the delayed test, even after the correction for attenuation.

b. Lessons. In order to estimate the reliability of the tests in ordering the 19 lessons, the following procedures were followed at each testing point. First, an average score was computed for each item by averaging over students. Next, an average score for each lesson was computed by averaging over the item scores for each lesson in a given set of items. This provided six estimates (one for each item set) of lesson difficulty. These were analyzed as a 6 (item sets) by 19 (lessons) score matrix in order to provide an estimate of reliability at that particular testing point. These estimates can be found, in parentheses, along the principal diagonal of Table 4. The reliabilities are again associated with the standard deviations, but the degree of reliability is not as high as it was for the individual items.

The correlations between testing points can be found above the principal diagonal in Table 4. The lesson scores used in this analysis were computed by averaging over all item scores for a given lesson at a given testing point.

TABLE 4

Reliabilities and Intercorrelations for
Seven Testing Points: Lessons

Testing Points	Testing Points 1	2	3	4	5	6	7	S.D.
1	(.71)	-.04	.19	.21	.39	.47	.36	.15
2		(.61)	.61	.09	.25	.15	.29	.06
3			(.34)	.50	.72	.68	.68	.04
4				(.63)	.83	.78	.49	.10
5					(.82)	.91	.72	.14
6						(.64)	.71	.13
7							(.72)	.14

These coefficients were not corrected for attenuation. The reliability coefficients, as computed here, are affected by "errors" in the sampling of both students and items, so the practical interpretation of such a correction would not be as obvious as it was in the previous analysis. It might be noted, also, that the internal consistency model used in computing these reliabilities has provided estimates that, in several cases, are far below the reliabilities which actually limit the intercorrelations.

The general pattern of correlations is similar to that found for the individual items: the post-tests fall into a fairly simplicial pattern, and the pre-test tends to be more highly related to the late post-tests than to the early post-tests. The pattern here does contain more irregularities, however.

The lesson means do not provide a very powerful method for locating weak items. In fact, the lesson means account for only 27%, 16%, 9%, 17%, 31%, 18%, and 24% of the variance of individual items for testing points 1 through 7, respectively. Any system of review that allocated effort purely on the basis of lesson or topic difficulty would allocate a good deal of time to items that did not need review and completely miss a number of items that did need review.

D. Discussion

The results of this study, in which the retention of electronic fundamentals was measured by means of criterion-referenced tests, do not differ appreciably from the results of previous studies in which retention was measured by means of norm-referenced tests. Retention in the interval between two weeks and two months, for example, was about nine percentage points higher in this study than it was for similar material in a study by Wickens, Stone, and Highland (1952). The loss in retention over a period of three years following the completion of the course was almost identical to that found by Williams and Whitmore (1959) for a test on "basic electronics." This correspondence is largely accidental, however, since it would be quite possible to design good norm-referenced tests which would provide measures of retention that range from one end of the scale to the other.

It was found that about half the gain in scores had been lost by the end of the course, and that this decline continued, at a diminishing rate, over several years on the job. This does not mean that the students "forgot" half of what they learned, however. Any index of retention is highly dependent upon the particular measurement technique that is used. In the current study, retention was measured by means of unaided recall, since it was felt that this would provide the best estimate of the information and skills that would actually be available for use on the job. Had some other technique been used, for example, recognition or relearning, the amount of measured loss would probably have been smaller.

An effort was made to use instructor ratings as a technique for discovering the topics which, if forgotten, would most adversely affect subsequent learning or performance on the job. It was found that there was very little agreement on these ratings. Since the agreement that was found was probably inflated to some extent by invalid stereotypes, caution would seem to be in order when using information of this kind as a basis for the design of training sequences.

It was found that the topics rated as being most relevant to the job were remembered somewhat better than the remaining topics, but that 35 percent of this "important" material had been forgotten by the technician on the job. This forgetting might be explained in part by ambiguities or unimportant particulars which, in this particular set of items, were inextricably confounded with a more basic set of principles that are remembered. On its face, however, the data suggest either an extreme heterogeneity of jobs, a lack of validity in the judgement of relevance, or a less than optimal level of performance by technicians on the job.

The programmed booklets used in this study had all been validated against immediate post-tests that covered two to three hours of instruction. These tests play a vital role in detecting deficiencies in the programs, but they do not provide an adequate means for controlling the proficiency of course graduates. The data from testing point 6 indicate that almost a third of the items are being missed by a majority of the graduates.

The obvious solution to this problem is some kind of review procedure (beyond the reviews that are currently being used in the course), but the problems involved in such an approach are more imposing than one might think. An extrapolation from the materials used in this study suggests that the course as a whole would comprise about 1500 separate training objectives. The objectives covered in the present sample were generally tested against a single "item," but a number of these "items" were composed of several relatively independent problems. Thus, a testing of the course as a whole would probably require in excess of 3000 hand-scored problems.

It is unlikely that the difficulties imposed by such a mammoth testing program could be ameliorated to any great extent by sampling from general topic areas, since the current study indicates that topic areas can account for no more than a small percentage of the variance in individual items. The most promising approach to simplification is through a culling of objectives which are of minor importance or which have already served their intended purpose, in spite of the difficulties encountered in trying to do this for the objectives covered in this study.

APPENDIX A

LIST OF PROGRAMS

1. Elements of Electrical Physics - Matter
2. Elements of Electrical Physics - Dynamics
3. Elements of Electrical Physics - Conductors, Resistors, Insulators
4. Electrical Calculations - Conversion of Electrical Units
5. Electrical Calculations - Work, Power, and Energy (Electrical)
6. D. C. Circuits - Parallel Circuits
7. Magnetic Theory - Magnetism
8. D. C. Meters - Meter Movements and Scales
9. D. C. Meters - Voltmeters
10. D. C. Meters - Multimeters
11. Electromagnetic Devices - Generators
12. A. C. Theory - Generation of a Sine Wave
13. Reactive Circuits - Inductance
14. Reactive Circuits - Transformers
15. Reactive Circuits - Capacitance
16. A. C. Circuit Characteristics
17. Parallel A. C. Circuits
18. Introduction to Vacuum Tubes
19. Voltage Regulation and VR Tubes

APPENDIX B

INSTRUCTIONS FOR RATING ITEMS

As part of a study on the retention of technical material, the Review Tests that are used with 19 of the programmed instruction booklets from Phase I were given at various points during the course. These tests contain a total of 218 items.

The forgetting of some parts of this material would be far less serious than the forgetting of other parts. Some parts, for example, are taught strictly to develop a given concept; once this concept has been mastered the original material can be forgotten without serious loss to the technician. Other parts, however, will be used on the job or in learning the materials that will be taught in subsequent courses; the technician should remember this material. I would like your help in identifying the materials that will be most needed at some later date.

Assume that you could provide review on an individual basis at the completion of the AFU(A) Course. In other words, you can test each student and, if he misses a given item, can provide him with a review of that material without imposing the same review on all other members of the class. Place a check (✓) in front of each item for which, if the student missed it, you would provide review. Do not assume that everyone will know the easy items.

Below the check mark, print a "J" if you feel that this information is really needed on the job. If you feel that a student who did not possess this information would be <u>seriously</u> hindered on the job, print a "+" after the "J". In making these judgements, remember that you should be concerned with the actual requirements of the job, not the intelligence of the student. If a technician does not know the capital of the United States you would probably not want him working on your plane, but he does not need this information to do a good job.

If you feel that a given item reflects information that the student will really need if he is going to learn the material that will be taught at some time <u>after</u> the completion of the AFU(A) Course, print a "T" under either the check or the "J".

You may do this rating at your leisure, but please do not discuss the job with other raters. The following table summarizes the codes:

If a given item is missed:

✓	It should be reviewed at the end of AFU(A).
J	The technician will be hindered on the job.
J+	The technician will be <u>seriously</u> hindered on the job.
T	The technician will be hindered in the courses that follow AFU(A).

APPENDIX C

INSTRUCTORS RATINGS OF ITEMS

In order to reduce the amount of rating required of each judge, each of the six item sets was submitted to a different group of nine judges. Each group consisted of three instructors from each of the three phases of the AFU course. The reliabilities of the ratings were estimated by calculating separate reliabilities for each group and then computing an average by means of Fisher's z, across the six groups. For the ratings of relevance to the job, a J was scored as 1 and a J+, as 2. These reliabilities, together with average correlations between judges and average percentages of endorsement, can be found in Table 5.

TABLE 5

Ratings of Relevance to Job or to Subsequent Training

Rating	Reliability 9 Judges	Avg. Correlation	Avg. % Endorsed
Needs Review	.50	.12	48
Needed on Job	.81	.31	43
Needed for Training	.53	.12	59
Corrected Review	.65	.18	55
Summary	.77	.28	

It was assumed that the Review response would serve as a totally redundant summary for the Job and Training responses, but, as can be seen in Table 5, this was not the case. More items were endorsed as being required for subsequent training than were checked as needing review, and, even though it cannot be seen from the summary data, some items were endorsed as J+ without being checked for Review. In order to obtain a more consistent index, a Corrected Review response was created by assuming a positive response to each item that was checked for Review, or was endorsed as J+, or was endorsed as both J and T. A final Summary index was computed by assigning a value of 1 to Review checks, J endorsements, and T endorsements, a value of 2 to J+ endorsements, and then summing across categories.

There were just about as many J+ endorsements as J endorsements, but several judges confined themselves exclusively to one or the other. As a result, the use of three response categories on this variable did not have as great an effect on reliability as one might think. If all positive responses, whether J or J+, are scored the same, the reliability is still .78.

APPENDIX D

THE ROLE OF COMPUTATIONS IN ELECTRONIC MAINTENANCE

The early courses in electronics were designed to provide the technician with the skills that he would need in troubleshooting a piece of equipment with little information beyond that provided on a sparsely annotated schematic diagram. Among these skills was the ability to compute the readings that should be obtained from various test points, and to determine the way in which a signal would be influenced by various kinds of malfunctions. Over the years, however, more and more of this information that was once available only through computations has been provided directly in the various documents available to the technician. Although time devoted to instruction and drill in computations has been reduced, the general outline of the electronics courses does not appear to have been affected to any great extent by these changes.

The computational skills being taught in the current courses are generally viewed as "enabling" skills. In other words, they are taught so as to facilitate the learning of other, more job oriented skills. This view is reflected in the fact that, of the several dozen computational items tested, only two were included in the set of 44 items that the instructors considered most relevant to the job. Nevertheless, almost all of the computational items were endorsed by some instructors. The 13 items in the "basic" set were endorsed, on the average, by five out of nine judges. An additional 20 items, related specifically to a-c circuit theory, were endorsed, on the average, by two out of nine judges.

Because of this belief by some judges that the computations would be needed on the job, an effort was made to estimate the level of performance that could be expected on tasks that required such computations. One of the first problems was to identify these tasks, and this proved to be rather difficult. The most frequently mentioned tasks included such things as the maintenance of transient aircraft for which the usual maintenance information was not locally available, or the modification of circuits when, because of some emergency condition, the required replacement parts were not available; but no one was particularly happy with these examples. In any case, it appeared that most of the tasks would require the successful performance of, not one, but several computations.

In order to obtain data that could be used to estimate probability of success for tasks requiring various numbers of computations, the 13 basic computational problems were abstracted from the regular tests and administered to a new sample of 21, high-experience, AVIB students. The average score on these items was 61%, a somewhat higher score than that made by similar students in the original sample. For purposes of the present analysis, however, the tests were rescored, counting each problem as a separate item and giving no credit for partial solutions. This resulted in a total of 23 items, almost half of which required no more than the proper manipulation of metric prefixes. Each student's score was expressed as a percentage, and these percentages were raised to successive powers in order to estimate probability of success for various numbers of computations. These estimates were then averaged over the 21 students. It was found that the probabilities of success for tasks requiring from 1 through 6 computations were 63%, 44%, 33%, 26%, 21%, and 17%, respectively.

Even though these figures indicate a fairly low probability of success, they represent a very inflated estimate of the proficiency one might actually expect to find on the job. These estimates are based on relatively easy computations, whereas most of the tasks suggested in the interviews would require the more difficult computations associated with a-c circuits and amplifying devices. It was found that the AVIB students had a 26% probability of success for trigonometric computations of the kind required for a-c circuit work, and that this probability dropped to 12% when these skills were applied to representative a-c circuit problems. Had problems of this kind been used as a basis for the estimates, one would have concluded that an experienced "A" school graduate has essentially no chance of working his way through a task that requires as many as three or four of these computations.

As was noted earlier, most judges viewed the computational skills as an aid to further training. In fact, 21 of the computational items were endorsed by at least seven out of nine judges as being "really needed" in the courses that follow the AFU course. The average score on these 21 items at the 6th testing point (the end of the AFU course) was 61%. If the judges are correct in stating that a student will be hindered in subsequent courses if he does not possess these skills, then a number of students are being hindered. It might be profitable to provide the poorer students with additional practice in computation, or, alternatively, to investigate training procedures that are not so vitally dependent upon those computational skills.

REFERENCES

1. Glaser, R. Instructional technology and the measurement of learning outcomes, American Psychologist. 1963, 18, 519-521.

2. Longo, A. A. and Mayo, G. D. Comparison of conventional and programmed instruction in teaching avionics fundamentals. San Diego: U. S. Naval Personnel Research Activity (Technical Bulletin STB 66-16), 1965.

3. Mayo, G. D. and Longo, A. A. Comparison of conventional and programmed instruction in teaching elements of electrical physics. Memphis: Naval Air Technical Training Command, 1964.

4. Wickens, D. D., Stone, G. R., and Highland, R. W. A study of the retention of electronics fundamentals during basic radar mechanic training. Lackland AFB: Human Resources Research Center (Research Bulletin 52-36), 1952.

5. Williams, W. L., Jr. and Whitmore, P. G., Jr. The development and use of a performance test as a basis for comparing technicians with and without field experience. Washington, D. C.: Human Resources Research Office (Technical Report 52), 1959.

DISTRIBUTION LIST

Chief of Naval Personnel (Pers-A3) (20) (Pers-C) (3)

Chief of Naval Operations (5)

Chief of Naval Air Training

Chief of Naval Air Technical Training (60)

Chief Administrative Services and Data Support Division,
U. S. Military Academy, West Point, New York 10996

Defense Documentaiton Center (20)

Naval Personnel Research and Development Laboratory (3)

Office of Naval Research (Code 458)

Office of Naval Research Branch Office, London (15)

Office of Naval Research Branch Office, Pasadena, California

U. S. Naval Academy, Annapolis, Maryland

U. S. Air Force Academy, Colorado Springs, Colorado

Air Training Command, Randolph Air Force Base, Texas

Air Force Human Resources Laboratory, Brooks AF Base, Texas

U. S. Army Behavioral Science Research Laboratory, Arlington, Virginia

Interagency Committee on Manpower Research (2)

Office of Secretary of Defense (MMRC)